An Insight Into Heavy-Tailed Distribution

Annapurna Ravi †
Ferry Butar Butar ‡

ABSTRACT

The heavy-tailed distribution provides a much better fit to financial data than the normal distribution. Modeling heavy-tailed distributions is done by resorting to stable distribution. The parameters of stock-market are estimated by the MLE. To check the superiority of stable over normal distribution, we used both graphical methods as well as test statistics for normal and for stable distribution.

INTRODUCTION

Over decades, development and modeling of financial concepts was based on the assumption that the financial data was distributed normally (i.e. the data was supposed to possess Gaussian distribution). That was the case with option pricing, risk analysis, etc. But the huge losses incurred (for example, cases of Barings or Daiwa and huge stock market downtrends) made the analysts to go back and locate the roots of the mis-analysis. It was Mandelbrot (1963), who found that the data was highly skewed and also had huge tails and recognized that these were not the characteristics of Gaussian distribution. He finally concluded that the financial data have a non-Gaussian distribution with huge tails. This distribution with huge tails was named as "Heavy-tailed distribution". They are also known as Power-law distribution (since they have power-law decay), fat-tailed, long-tailed distribution. These distributions are hyperbolic in nature and are highly skewed. An important thing to remember is the skewness is only a possibility. Even the distributions which are not skewed but have huge tails are said be heavy-tailed distributions. Examples of heavy-tailed distributions are the Pareto, Levy, log-gamma distributions. The most interesting feature and the feature that made these distributions popular is that these distributions can accommodate extreme values (i.e. extremely large values or extremely small values). In these distributions, the data will have power-law decay instead of the usual exponential decay, which occurs in Gaussian distributions. The existence of heavy-tailed distributions is not limited to finance. Even most of the data in economics, geology, climatology, signal processing, insurance, environmetrics do have heavy-tailed distributions. See Rachev (2003), Adler et.al., (1997), Embrechts et.al., (1953), Coles (2001), Barry (1983), Zolotarev (1986), Press (1972), and among others. The heavy-tailed distributions are defined as

$$P(X > x) \sim m x^{-\alpha} \quad as\ x \to \infty, and\ 0 < \alpha < 2, \tag{1.1}$$

where X is a random variable, α is a shape parameter, and m is a location parameter.

Heavy-Tailed Distribution

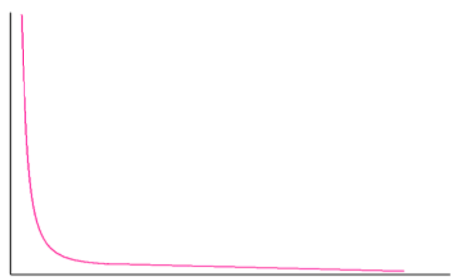
Identifying Heavy-tailed distribution

Over the course of time, a large number of methods was developed to identify whether a given data set has a heavy-tailed distribution or not. This is very important since this conclusion makes the data to be considered either Gaussian or Non-Gaussian. This classification can be done either mathematically or graphically. If the distribution of a given data is known and if it can be expressed in the form of equation (1.1), then the data is said to have heavy-tailed distributions. But a problem arises when the distribution of a given data is not known. Then the identification can be done graphically. There exist a lot of graphical methods to carry out the identification process. Some of them are normal probability plots, box plots, and CCDF (Complementary Cumulative Distribution Function) test.

Graphical methods to identify heavy-tailed distribution

(a)Plot the given data and check if it exhibits hyperbolic nature. If yes, then the data have heavy-tailed distribution. See figure 1 below.

Figure 1: Hyperbolic distribution



(b) Normal Probability plots: Normal probability plot also known as pp-plot is one of the graphical means for determining normality of the given data. Normal probability plots provide the information about the outliers in a given data and skewness of the data graphically. In the normal probability plots, the values of the given dependent variable (arranged in ascending order) are plotted against the percentiles of a normal distribution. If the graph is linear then the data are said to be normal. Given below are different cases of a probability plot. The

Figure 2 shows the probability plot for left-skewed data. The plot starts below the line, crosses the line and then ends below the line. Figure 3 shows the probability plot for right-skewed data. It starts above the line, crosses the line and then again ends above the line. Figure 4 shows the probability plot for data having symmetric heavy-tailed distributions. The plot starts below the line, crosses the line in the lower end stays above the line. When it reaches to about the middle of the line, the plot crosses the line and stays below. Finally it ends above the line. Note that if the data possess heavy-tailed distribution and is not symmetric, the basic outline of the plot remains the same except that the plot do not cross the line exactly in the middle, similarly figure 5 show probability plot of symmetric light-tailed data.

Figure 2: Normal probability plot for Left-Skewed data

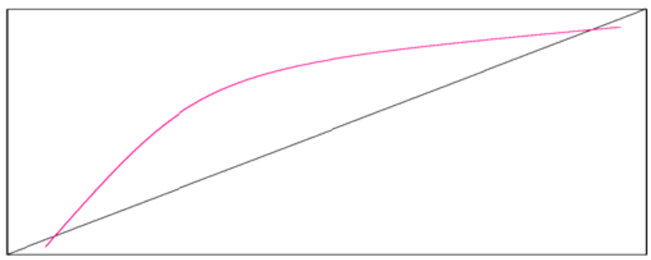


Figure 3: Normal probability plot for Right-Skewed Data

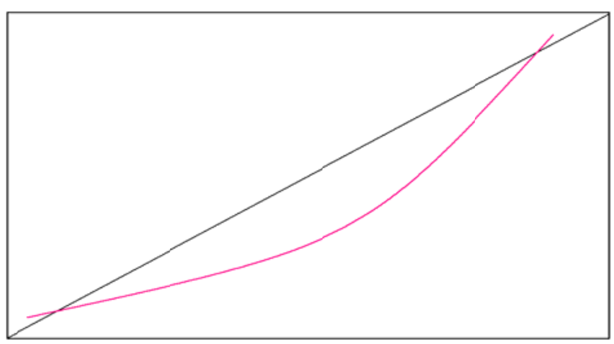


Figure 4: Normal probability plot Symmetric Heavy-tailed Data

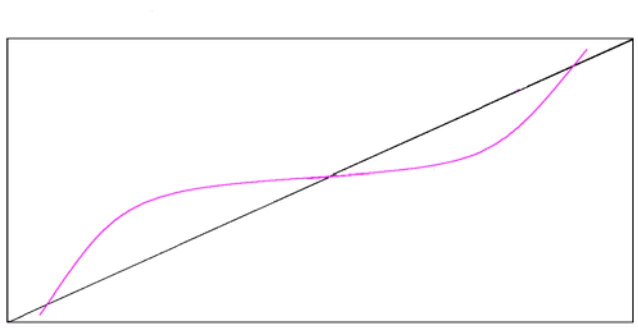


Figure 5: Normal probability plot for Symmetric Light-tailed Data

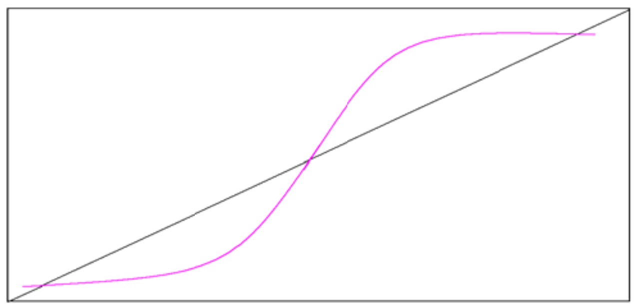
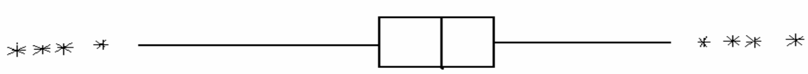


Figure 5 shows the normal probability plot for data having symmetric light-tailed distribution. The plot starts above the line, crosses the line in the lower end, stays below the line until it reaches the mid point of the line. When it reaches to about the middle of the line, the plot crosses the line and stays above. Finally it ends below the line.

(c) Box plot: Box-plot is a graphical way to represent data in terms of quartiles. Along with these, it also shows the lower limit, upper limit and any possible outliers in the given data. Any outliers in the given data are shown as stars, and are located away from the rectangular region. If the box plot for the given data has outliers on both sides and has tails longer than the length of the box, then the data is said to have heavy-tailed distribution. Example of such box plot is shown below.

Figure 6: Box plot for heavy-tailed distribution



Modeling Heavy-tailed distribution

Heavy-tailed distributions can be modeled by any of the following distributions such as Stable, Student's t, hyperbolic, Normal inverse Gaussian or truncated stable distributions. In this paper, stable distributions are considered to model the given data from a heavy-tailed distribution. The main reason for the selection of stable distribution is they are the only distributions supported by the generalized central limit theorem which are leptokurtic (a distribution is said to be leptokurtic if its kurtosis is less than 3). Most of the financial data (in general case empirical data) discussed above follow heavy-tailed distribution and in most cases are asymmetric, and so cannot be modeled by Gaussian distributions. Therefore, stable distributions are the only alternative.

Stable Distribution

Consider the variables X_1, X_2, X_3..., X_n that are independent, identically distributed variables. If $X_1 + X_2 + \cdots + X_n \xrightarrow{d} a_n X + b_n$, where n is a positive integer, $a_n > 0$, and b_n is a constant, then X_1, X_2, X_3... X_n are said to have Stable distribution. In the above equation, a_n usually takes the form of $n^{1/\alpha}$. Detailed discussion about α is given below. If n independent random variables have stable distribution and same index α, are added, the resulting distribution is again a stable distribution with index α. However, this condition is not satisfied when the variables have different index α, i.e. they exhibit invariance property of α. Since there is no closed form expression for the densities of stable distributions, it is described by a characteristic function of $S_{\alpha,\beta}(\delta, c)$, which is the Inverse Fourier Transform of the PDF which is given as follows:

$$\int_{-\infty}^{\infty} e^{itx}\, dH(x) = \begin{cases} \exp\left[-c^{\alpha}|t|^{\alpha}\left(1 - i\beta sign(t) \tan\left(\frac{\pi\alpha}{2}\right)\right) + i\delta t\right] & \alpha \neq 1 \\ \exp\left[-c|t|\left(1 - i\beta\left(\frac{2}{\pi}\right) sign(t)\, ln|t|\right) + i\delta t\right] & \alpha = 1 \end{cases}$$

(1.2)

where H is the Distribution Function, and α where $0 < \alpha \leq 2$ is the Characteristic Exponent or index of stable distribution. In literature there are different notations for the four parameters of a stable distribution i.e. α, β, δ, and c whose details are presented in the table below.

Table 1: Stable distribution parameters

parameter	Name	Possible values
α	Index of Stability, Tail index, Tail Exponent	$0 < \alpha \leq 2$
β	Skewness Parameter	$-1 \leq \beta \leq 1$
δ	Location	$\delta \in R$
c	Scale	$c > 0$

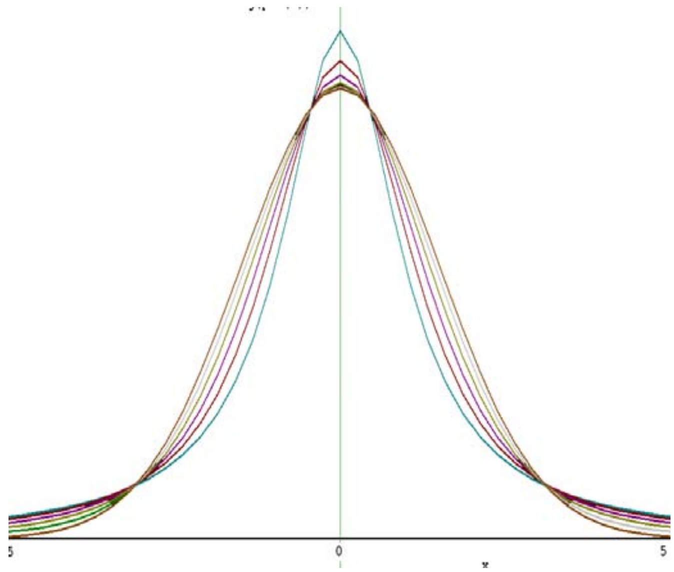
α determines the type of distribution and the length of the tails. The length of the tails increases as the value of α decrease from 2 to 0. If α=2, the distribution is normal, and if α=1, the distribution is Cauchy. If α is in the range of 0 to 2, then the distribution is a Stable distribution. Higher order moments such as variance and others exist only when α = 2. If, for a distribution, α is less than 2, then the distribution is leptokurtic and has fat tails, and the variance is infinite. This seems to be inappropriate. But in the distributions with infinite variations, one of the summands contributes the most to the sum of variables. This can be perfectly applied to a case, when there is a probability for large deviations in a single variable, while this type of probability can be ruled out, or is minimum in the case of remaining variables. This is a perfect fit for situations that occur frequently in Stock markets, financial institutions, earthquakes, etc. As discussed, the existence of mean and variance depends on α. When α equals 2 both mean and variance exist. But when α is in the range of 1 and 2, the mean exists while variance becomes infinite. Even in such cases, the variance of the distribution can be measured. Since the mean exists, the absolute mean deviation can be calculated and can be used as a measure of variance of the distribution. When α < 1, both mean and variance are infinite. In some cases, the tail index exponent α helps in finding the estimates of the remaining variables. According to Mandelbrot (1963), "when α is greater than 1, the location parameter δ is equal to the mean of the distribution". β determines the skewness of the distribution. If the value of β ranges from -1 to 0, then the distribution is left-skewed. If its value is equal to zero, then the distribution is symmetric. Instead if the value of β is greater than zero but less than one, then the distribution is right-skewed. When α starts approaching 2, the distribution starts becoming Gaussian, irrespective of the value of β. Note that β is zero in case of Gaussian distribution or symmetric stable distribution. The parameter δ as described above is location parameter, while c is the scale parameter.

Behavior of Stable distributions under different parameter conditions

Case 1: Effect of α on Symmetric stable distributions:

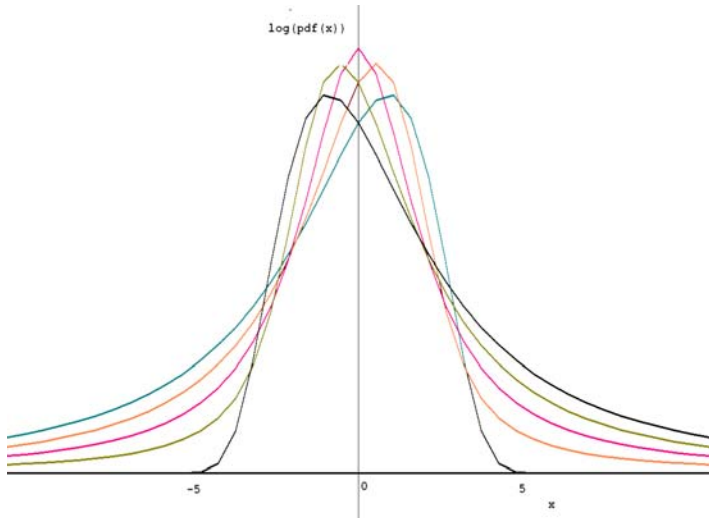
Shown below in the figure 7 is the picture to show dependence of the distribution on α when the distribution is symmetric. In this distribution, α ranges from 1 to 2, β = 0, c=1 and δ =0. It can be seen that when α=2, the curve converges and exhibits Gaussian behavior. But when it starts decreasing towards 1, the time taken to converge increases. In figure 2.7 for α = 1.0, 1.2, 1.4, 1.6, 1.8. 2.0 the color of the graph is blue, red, pink, grey, green, and orange, respectively.

Figure 7: Symmetric stable distribution for 1.0 ≤ α ≤ 2.0, β=0



Case 2: (i) Effect of α on Asymmetric stable distributions (right-skewed distribution).
(ii) Effect of α on Asymmetric stable distributions (left-skewed distribution).
See Ravi (2005).
Case 3: Effect of β on stable distributions: Shown below in figure 8 is the effect of β on the stable distribution. When $\beta < 0$, (in this case, β=-1, -0.5 as shown by black and green curves) the distribution is left-skewed. When $\beta = 0$, (in this case as shown by the pink curve) the stable distribution is symmetric. It reaches its maximum value at zero. And when $\beta > 0$, (in this case, β=0.5, 1 as shown by red and blue curves) the stable distribution is right-skewed. In figure 8 describes β = -1.0, -0.5, 0, 0.5, 1.0 for color of the graph is black, green, pink, red and blue, respectively. See Ravi (2005) for alpha = 1.5 and 2.0.

Figure 8: Stable distributions for $\alpha=1.0$ and $-1 \leq \beta \leq 1$

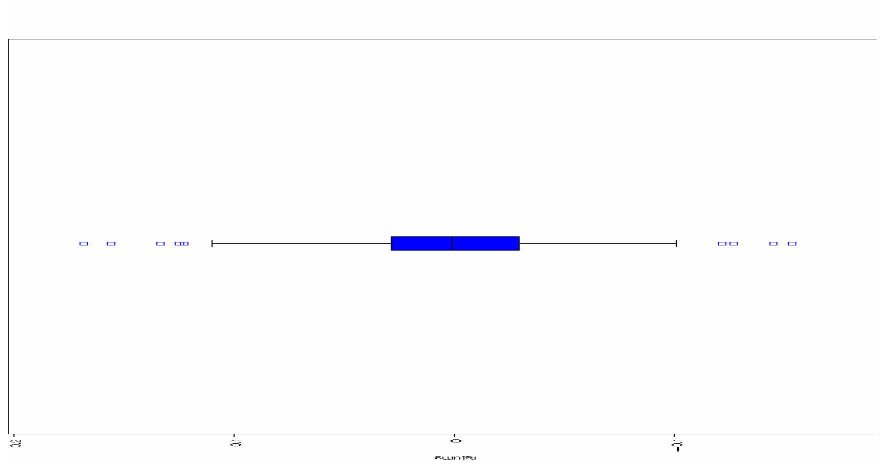


Testing For Normality

In this paper, the data are collected from the closing prices of Dow-Jones Industrial average from Jan 1 1984 to Dec 31, 1997. The reason for the selected large span of time is that there were a lot of studies which were based on the assumption that when data comes from the stock market and is large, it exhibits properties of normal distribution. The main aim of this paper is to show that for such data from stock-market, heavy-tailed distributions are better fit when compared to that of normal distribution. Several graphical tests such as box-plot, probability plot and also tests such as the Shapiro-Wilk test, Kolmogorov-Smirnov test, Anderson-darling test and the Cramer-Von test can be carried out to prove that the given data is normally distributed or not.

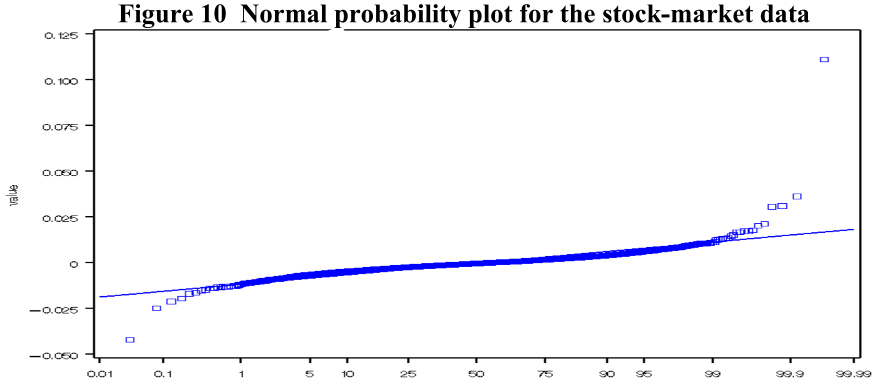
(1) Box plot: In figure 9, the box-plot for the data is shown. There are outliers on both sides of the rectangle. The length of the lines attached to rectangle is long when compared to the length of the rectangle indicating that the distribution has heavier tails. Therefore when this box plot is compared with the prototype box-plot, it can be concluded that the data is not from a normal distribution.

Figure 9: Box-plot for the stock-market data



(2) Normal Probability plot: The normal probability plot for the selected stock-market data is shown in figure 10. Although, for 90% of time, the plot remains on the line, the following important points must be noted. The plot starts below the actual line. Then it crosses the line and stays above the line. In the middle of the plot, the plot again crosses the line and stays below the line. Immediately, it crosses the line and ends above the normal line. So the behavior of probability plots for the given stock-market data does not match with that of the probability plot from normal distribution. Therefore, we will conclude that the given stock-market does not have normal distribution based on its normal probability plot behavior.

Figure 10 Normal probability plot for the stock-market data



(3) Histogram: From the histogram (see figure 11), and fitting of the normal curve on the histogram, it can be concluded that the data is from normal distribution. This conclusion is different when compared to the conclusions from the box plot and normal probability plot.

To confirm the distribution of the data, whether it is normal or not, a series of statistical tests such as the Kolmogorov-Smirnov, Anderson-Darling, the Cramer-Von Mises test are performed.

Figure 11 Histogram for stock-market data

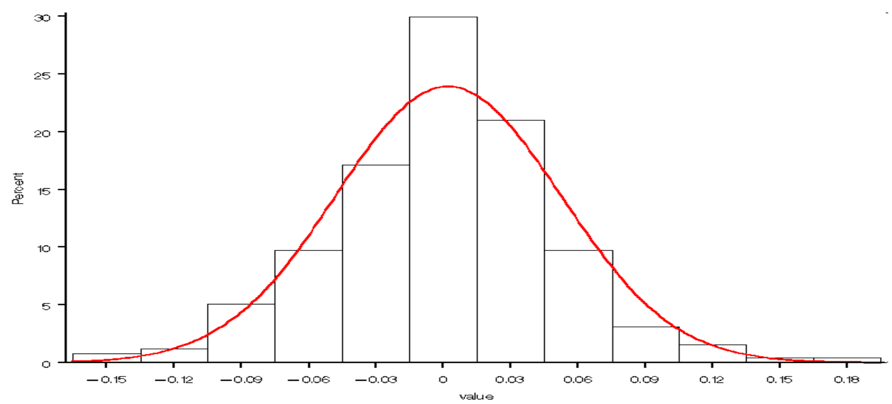


Table 2:Test statistics for testing normality

Test	Normal	Stable
Kolmogorov-Smirnov	D=0.10998; P < 0.01	D=.0544; P=0.25
Cramer-Von Mises	W-Sq=10.4064; P < 0.005	W-sq=0.1476; P=0.08
Anderson-Darling	A-Sq=60.5085; P<0.005	A-sq=0.8939; P=0.06

Table 2 shows the tests, their statistics and p-values. The Kolmogorov-Smirnov, Cramer-Von Mises, and the Anderson-Darling tests are performed on the given stock-market data. For the normal test, all p-values of all these tests are less than 0.05. Therefore it can be concluded based on results of all the tests, that the stock-market data is not normally distributed. For Stable test, all p-values are larger than 0.05 meaning that the stock-market is from stable distribution. Although histogram shows that data is bell shaped, since all the other tests conclude that it is not normally distributed, it is finally concluded that the data are not from a normal distribution.

Modeling Data From Heavy-Tailed Distributions

As discussed earlier, the data of closing prices of Dow-Jones industrial average is collected from stock-market for the period Jan 1 1984 – Dec 31, 1997. In some of his early works, Moore (1991) was able to prove that weekly changes in stock prices from New York Stock Exchange (NYSE) had normal distribution. But he considerably ignored the fact that the distribution had longer tails than compared to normal distribution. In the words of Teichmoeller (1971), "Stock prices do not appear to exhibit the properties which would indicate that stock price changes are represented by a simple mixture of normal distributions". To support these discussions, series of tests are carried out on the collected stock-market data. Although in section of box-plot, it has been proved that the data is not from the normal distribution, the fact that the data possess heavy-tailed distribution is to be established. This is done by carrying out the box-plot test and probability plot test. Again using box-plot of the stock-market data (see figure 9), we observe that the outliers are on both sides of the rectangle, therefore the data posses very heavy tails. Since the length of whiskers seems to be proportionate to the length of the rectangle, so it can be concluded that the data is almost symmetric possessing heavy-tails. Based on the probability plot shown in figure 10, the normal probability plot test establishes the fact that the given stock-market data is from symmetric heavy-tailed distribution.

Logarithms of the collected stock prices are calculated. The main reason for calculating logarithms of the stock prices is for a given price level of a stock, it is observed that the variability of everyday price changes is an increasing function. Taking logarithm would eliminate the price-level effect. And one more point to be noted is, the estimation procedures are never employed on raw data available from the market. Instead the returns of daily stock prices are calculated and then estimation procedures are employed on these returns. These returns are calculated by taking the logarithm of the ratio of previous closing price and the present closing price. The values of the parameters of stable distribution are estimated based on these stock price return values by the method of Maximum likelihood estimation (DuMouchel, 1973). Although the maximum likelihood estimation method is slower (although not the slowest) when compared to other algorithms or methods currently in use, it almost yields the accurate estimators. Using the maximum likelihood, we can either perform direct integration or use FFT (Fast Fourier Transform) method which is equally good. The most commonly used is direct integration due to certain limitations on FFT method. Here, data are simulated using direct integration. The simulation results are presented below in table 3. The Stable distribution is then modeled using these parameter values and the corresponding curve is plotted.

Table 3: Estimated values of Stable distribution parameters for given Stock-market data

Parameter	Estimated value
Tail Exponent(α)	1.5153
Skewness Parameter (β)	0.1609
Scale Parameter (c)	0.0021
Location Parameter(δ)	-0.0002

The parameters briefly explain the behavior of the stable distribution curve for the collected stock-market data. The tail exponent(α) value being 1.5, indicates that although tails exist, they are not that heavy since its value is close to 2 (when α equal to 2, stable distribution becomes Gaussian). It also explains that the first moment i.e. the mean exists for this data. The skewness parameter being 0.16 indicates that the data is skewed right. Also the location parameter indicates a slight deviation from the center. The stable distribution is then fitted to the data. Figure 12 depicts this. Although, it resembles the normal curve, it is not. This can be noted by observing the tails of the curve. They seem to be converging but really don't. Thus, it preserves the property of the non-convergence of tails in heavy-tailed distributions. The reason it resembles normal curve is due to its tail exponent value which is equal to 1.5 (see stable distribution section regarding the behavior of stable distributions). In figure 13, this fitted stable distribution curve is superimposed on the normal curve so as to have a better understanding of the difference between the two fitted curves.

Conclusions

The main aim of this paper is to show that the stable distributions provide a better fit to stock-market data when compared to normal distribution. In section 4, the Stable distribution is fitted for the given stock-market data. This can be done either graphically or by carrying out series of tests. Proceeding graphically, the stable distribution fit and the normal curve for the stock-market data are super-imposed on each other in figure 13. When compared, stable distribution provides better fit to the histogram than normal distribution. This can be explained either in terms of height or in most of the cases where it touches the tips of the bars of the histogram. Therefore graphically stable distributions are proved to provide much better fit when compared to normal distributions. But to prove it technically a series of tests are performed. Hence, by both graphical means and performing tests, it showed that a stable distribution provide much better fit than when compared to the fit provided by a normal distribution.

Figure 12 Stable distribution fitted to histogram of stock-market data

Figure 4.2 Comparing Fitted normal and stable distributions

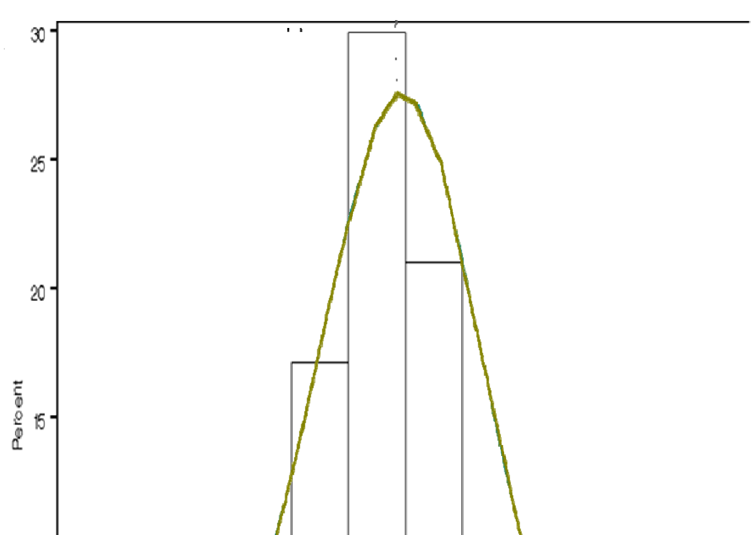
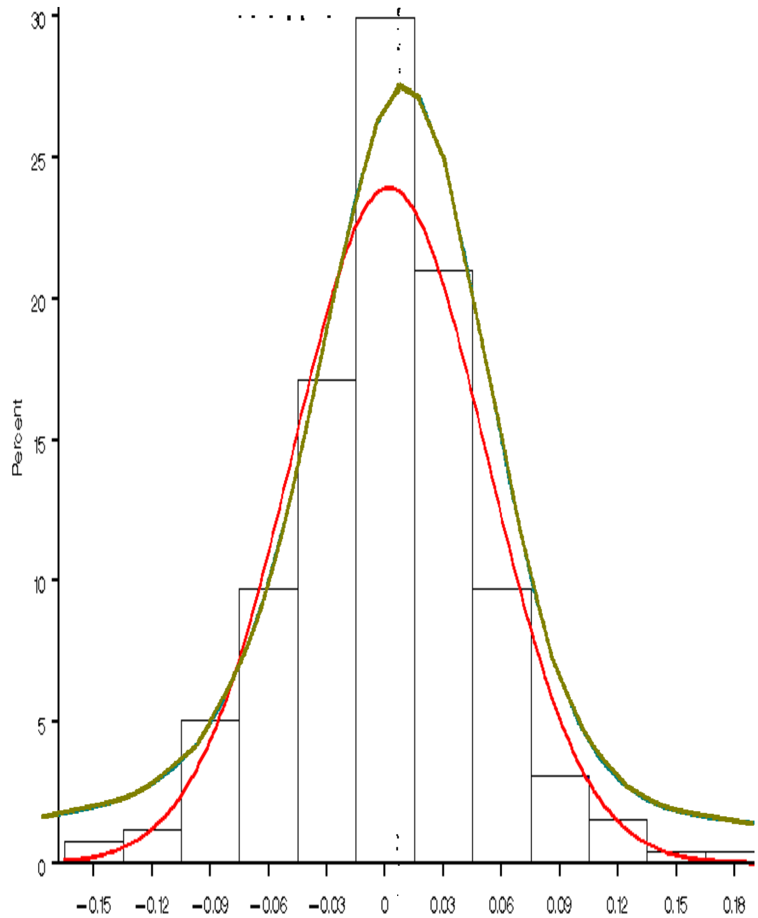


Figure 13: Comparing Fitted Normal and Stable Distribution



† Annapurna Ravi, Kendle International Inc.

‡ Ferry Butar Butar, PhD, Sam Houston State University, Texas

References

Adler, R., Feldman, R., and Taqqu, M., 1997, A practical guide to heavy tails: statistical techniques and applications, University of California Santa Barbara Dept of Statistics and Applied Probability.

Akgiray, V., and Booth, G.G., 1988, The Stable-law Model of Stock Returns, Journal of Business & Economic Statistics 6, 51-57.

Barry, C. A., 1983, Pareto distributions, Fairland, MD: International Co-operative Pub. House.
Coles, S., 2001, An introduction to statistical modeling of extreme values, London; New York: Springer.
DuMouchel, W.H., 1973, On the Asymptotic Normality of the Maximum-Likelihood Estimate when sampling from a Stable Distribution, The Annals of Statistics, 1, 948-957.
Embrechts, P., Kluppelberg, C., and Mikosch, T., 1953, Modeling extremal events for insurance and finance, London; New York: Springer.
Mandelbrot, B., 1963, The Variation of certain speculative prices, Journal of Business, 36, 393-419.
Moore, G. A., 1991, Crossing the Chasm, New York, Harper Publisher.
Press, S.J., 1972, Estimation in Univariate and Multivariate Stable distributions, Journal of American Statistical Association, 67, 842-846.
Rachev, S.T., 2003, Handbook of heavy tailed distributions in finance, Amsterdam; Boston: Elsevier.

Ravi, A., 2005, A non-Gaussian approach to achieve better fit for heavy-tailed distributions, Master thesis, Department of Mathematics and Statistics, Sam Houston State University, Huntsville, Texas.
Teichmoeller, J., 1971, A note on the distribution of the stock price changes, Journal of American Statistical Association, 66, 282-4.
Zolotarev, V.M., 1986, One-dimensional stable distributions, Providence, R.I.: American Mathematical Society.